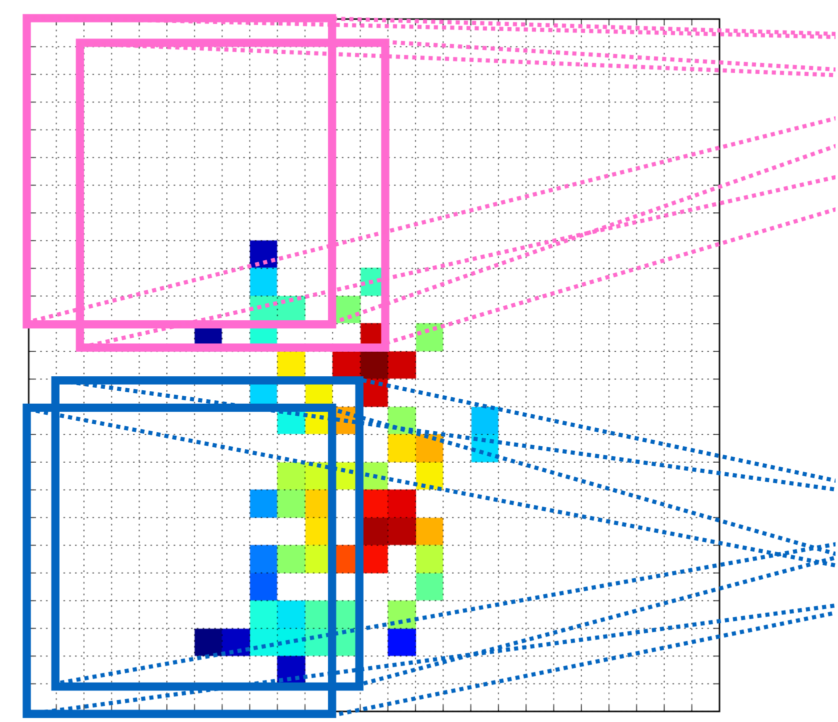
Deep Learning with the Largest Scientific Dataset

Modern machine learning for High Energy Physics

Benjamin Nachman

Lawrence Berkeley National Laboratory

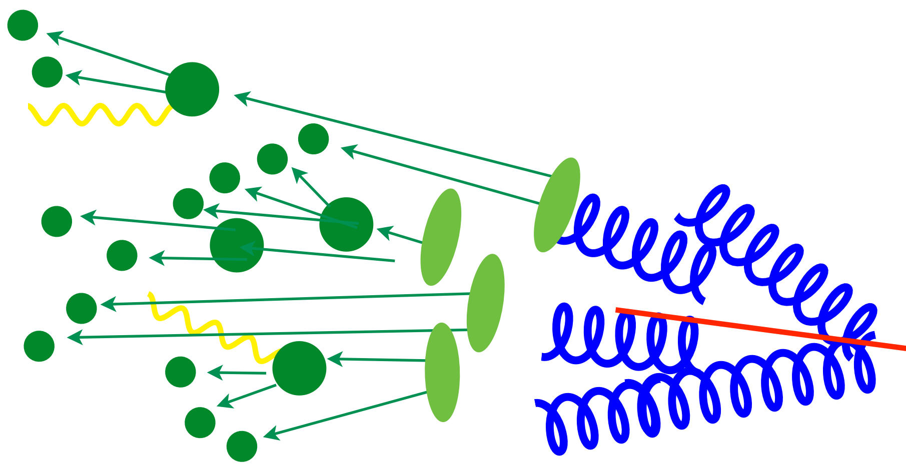


BERKELEY LAB

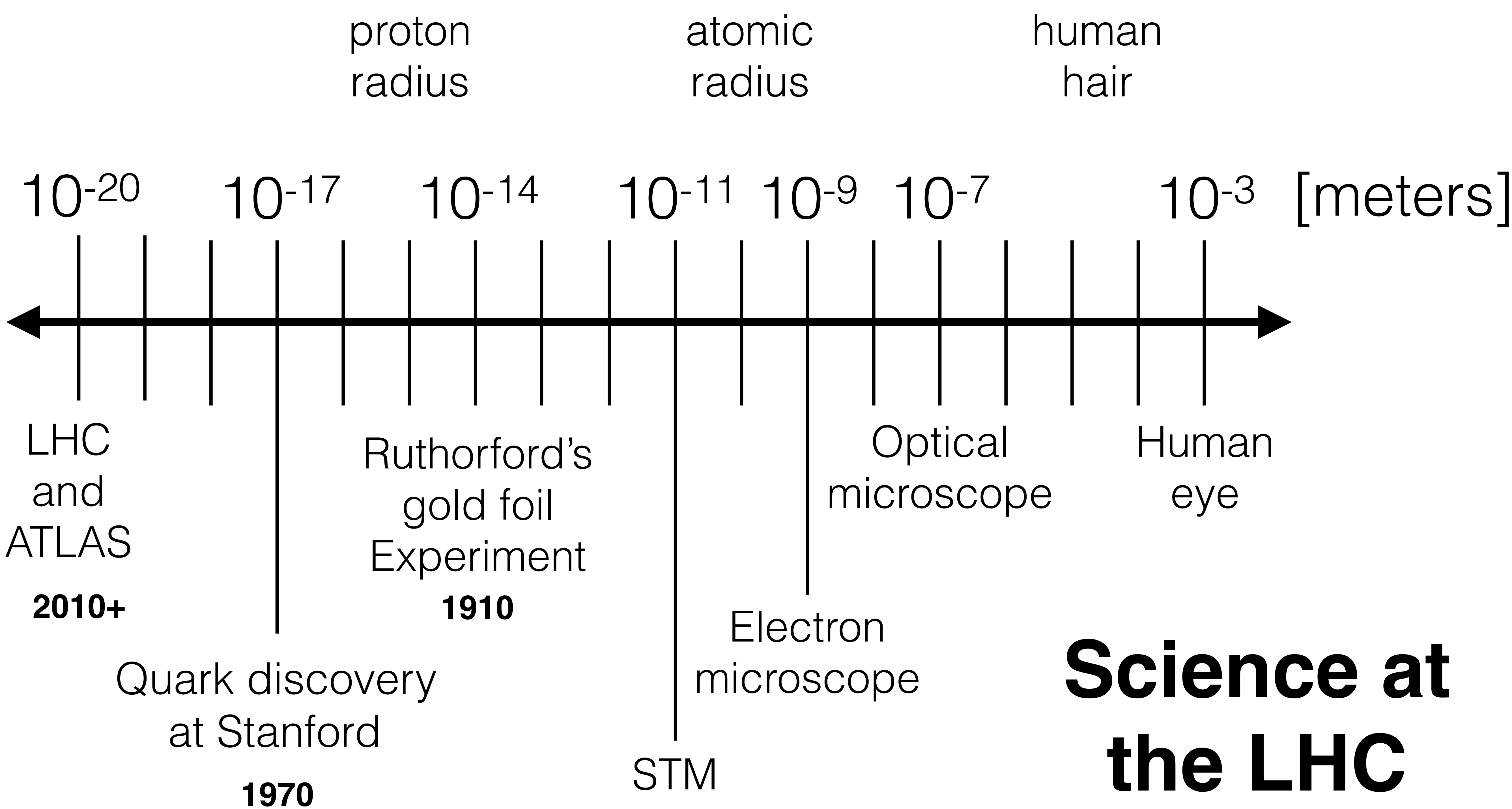
BERKELEY
EXPERIMENTAL
PARTICLE
PHYSICS

Outline

- A little bit of science
- Machine learning and high energy "jets"
- Applications of ML4Jets
  - ✦ CNN's for "pileup" noise
  - ✦ GANs for simulation
  - ✦ Weak supervision and learning from data



Goal: **We want to study the structure of the smallest building blocks of matter.** For this, we need the most powerful microscope ever built!



Science at the LHC

20 mile beam line and a 5 story detector

access to a new frontier!

10^{-20} 10^{-17} 1

LHC and ATLAS
2010+

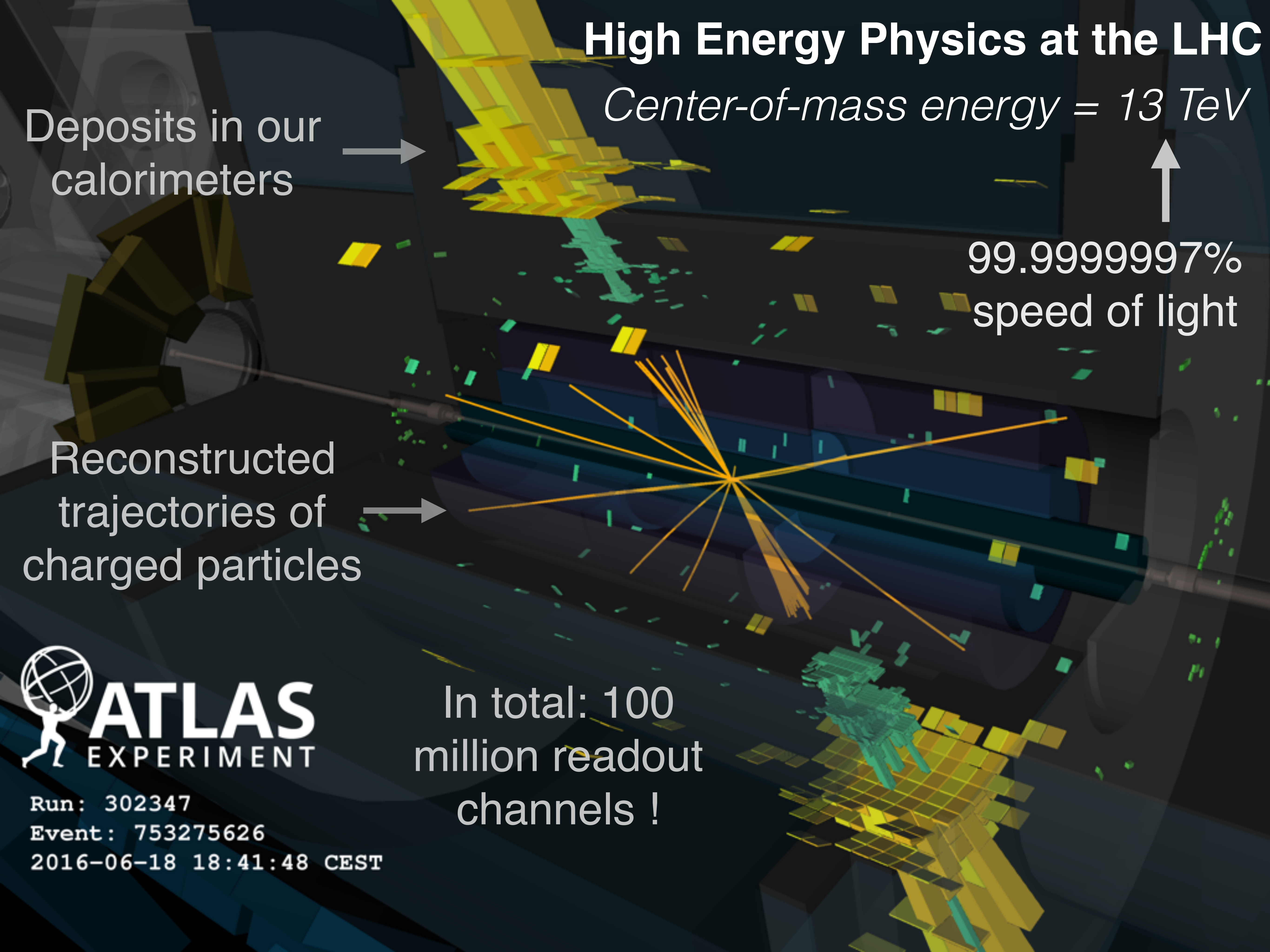
Quark discovery at Stanford
1970

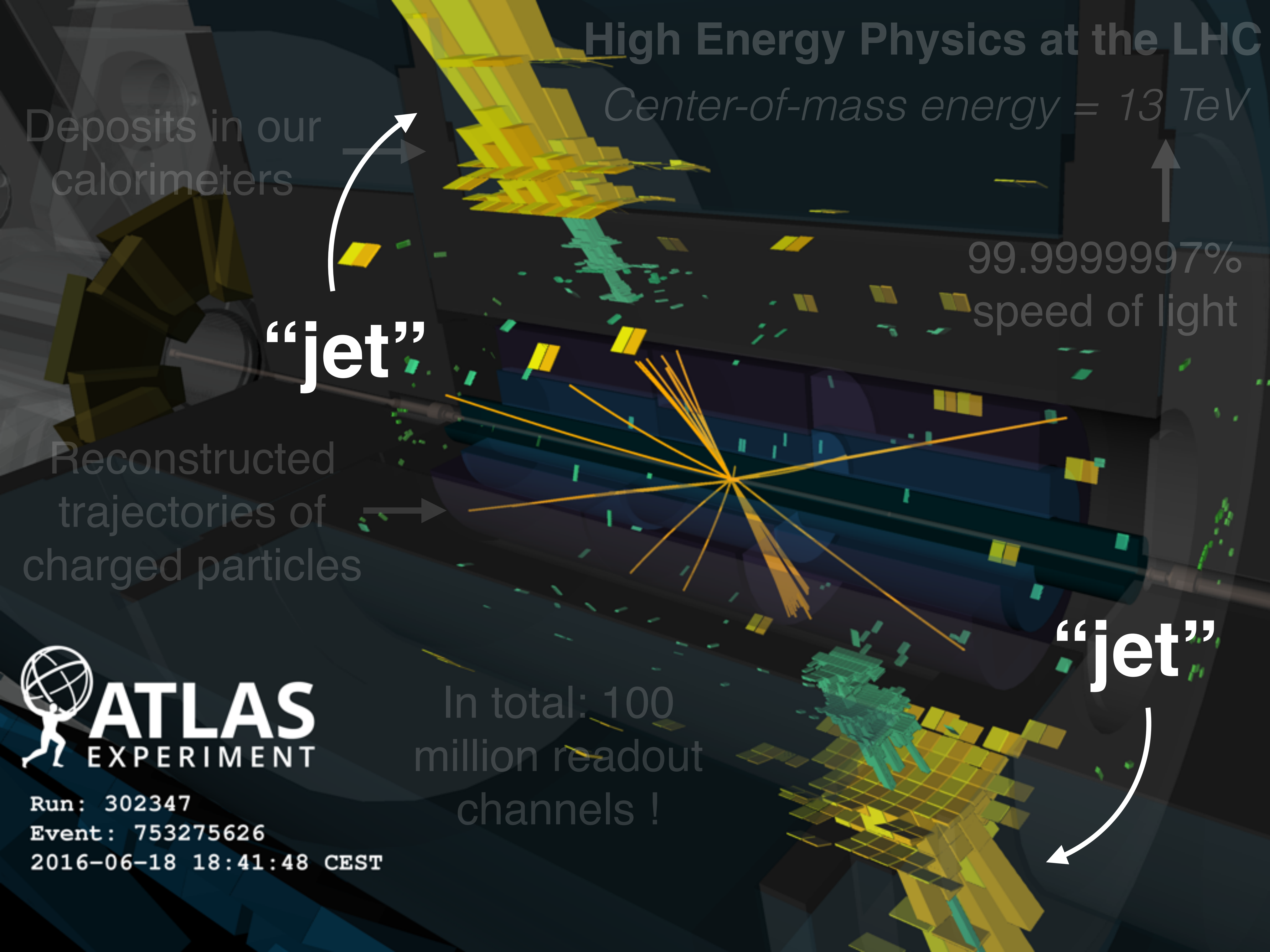
Ruthorford's gold foil Experiment
1910

STM

Ele
micr

Images: CERN

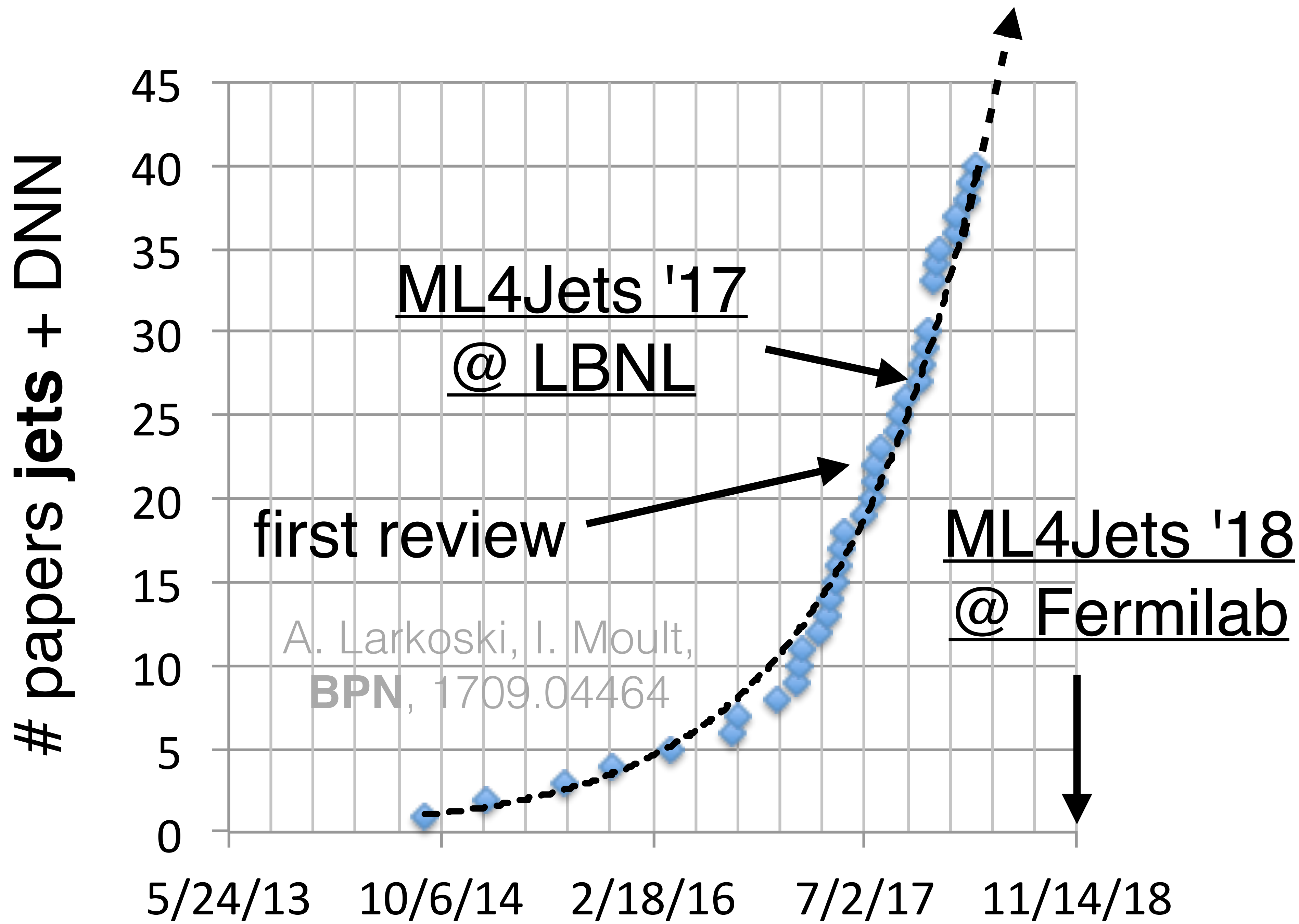

High Energy Physics at the LHC
Center-of-mass energy = 13 TeV
99.9999997% speed of light
Deposits in our calorimeters
Reconstructed trajectories of charged particles
In total: 100 million readout channels !
ATLAS EXPERIMENT
Run: 302347
Event: 753275626
2016-06-18 18:41:48 CEST


High Energy Physics at the LHC
Center-of-mass energy = 13 TeV
99.9999997%
speed of light
Deposits in our
calorimeters
"jet"
Reconstructed
trajectories of
charged particles
"jet"
In total: 100
million readout
channels !
ATLAS
EXPERIMENT
Run: 302347
Event: 753275626
2016-06-18 18:41:48 CEST

Jets with Machine Learning

classification, regression, generation, …

…this is a very active field of research!

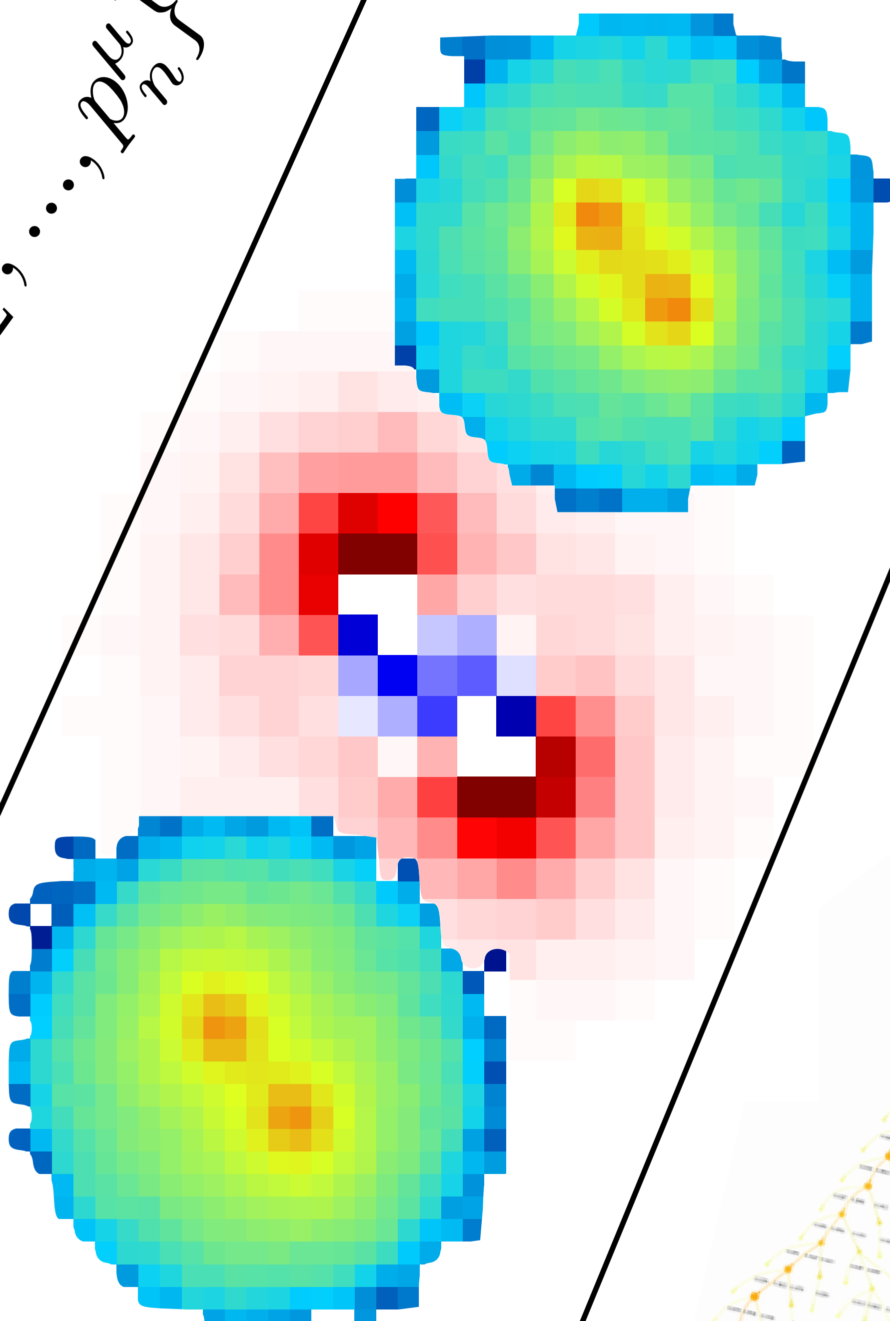


Organizing approaches: how do you think of jets?

Fixed sets

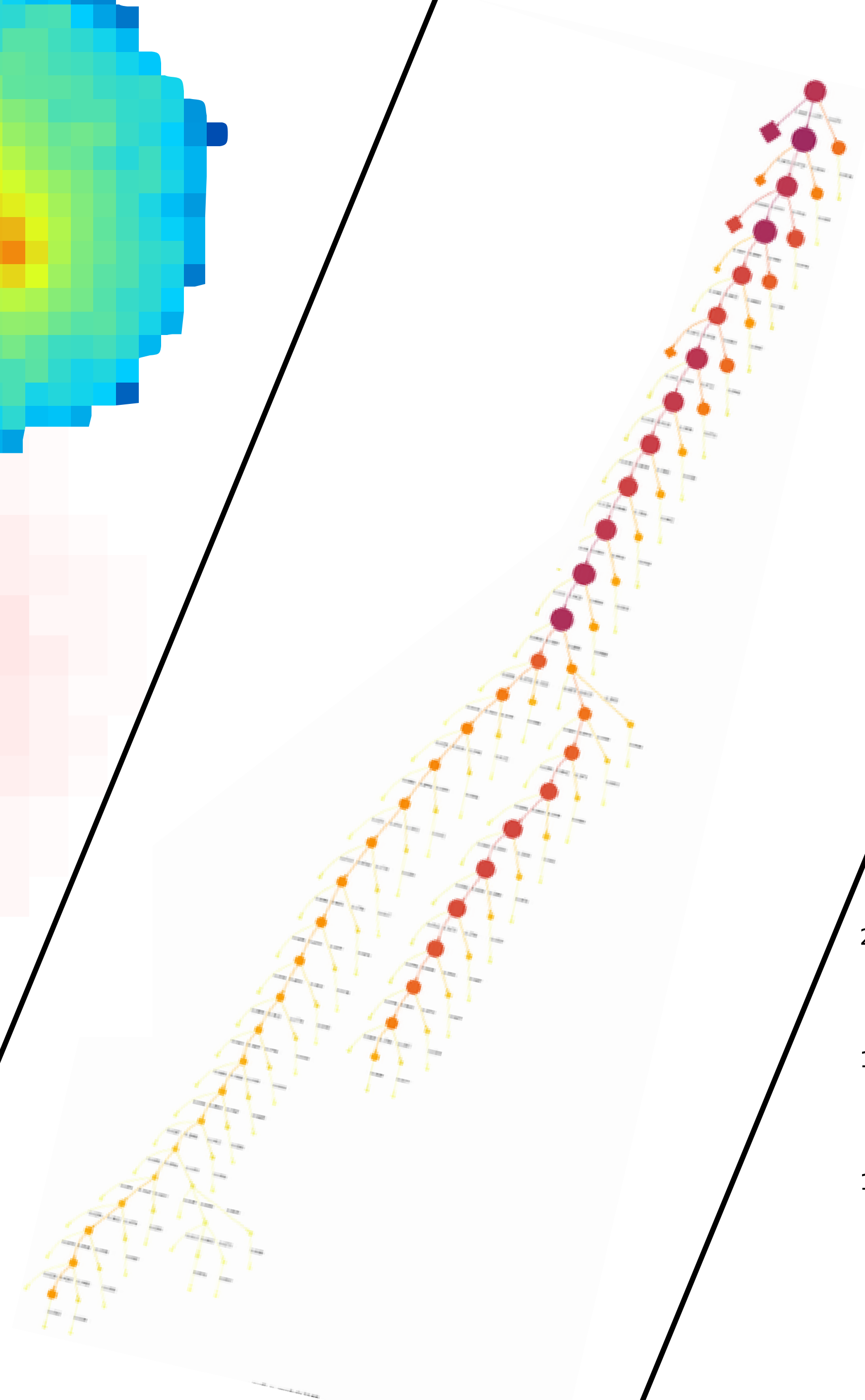
$$J = \{p_1^\mu, p_2^\mu, \ldots, p_n^\mu\}$$

Images



Jet Images: [1407.5675]

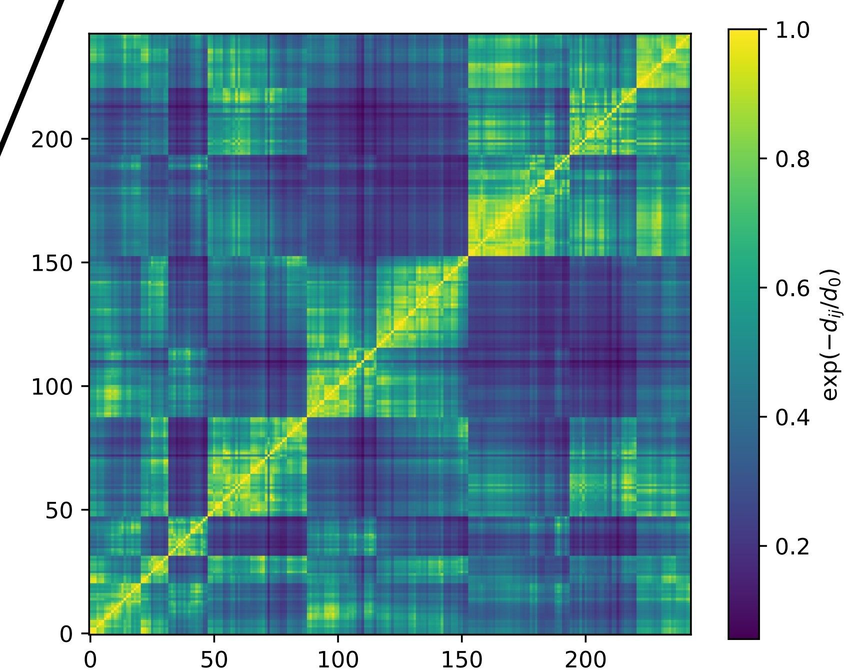
Sequences/ Trees



[1702.00748]

Graphs

[NIPS DLPS]

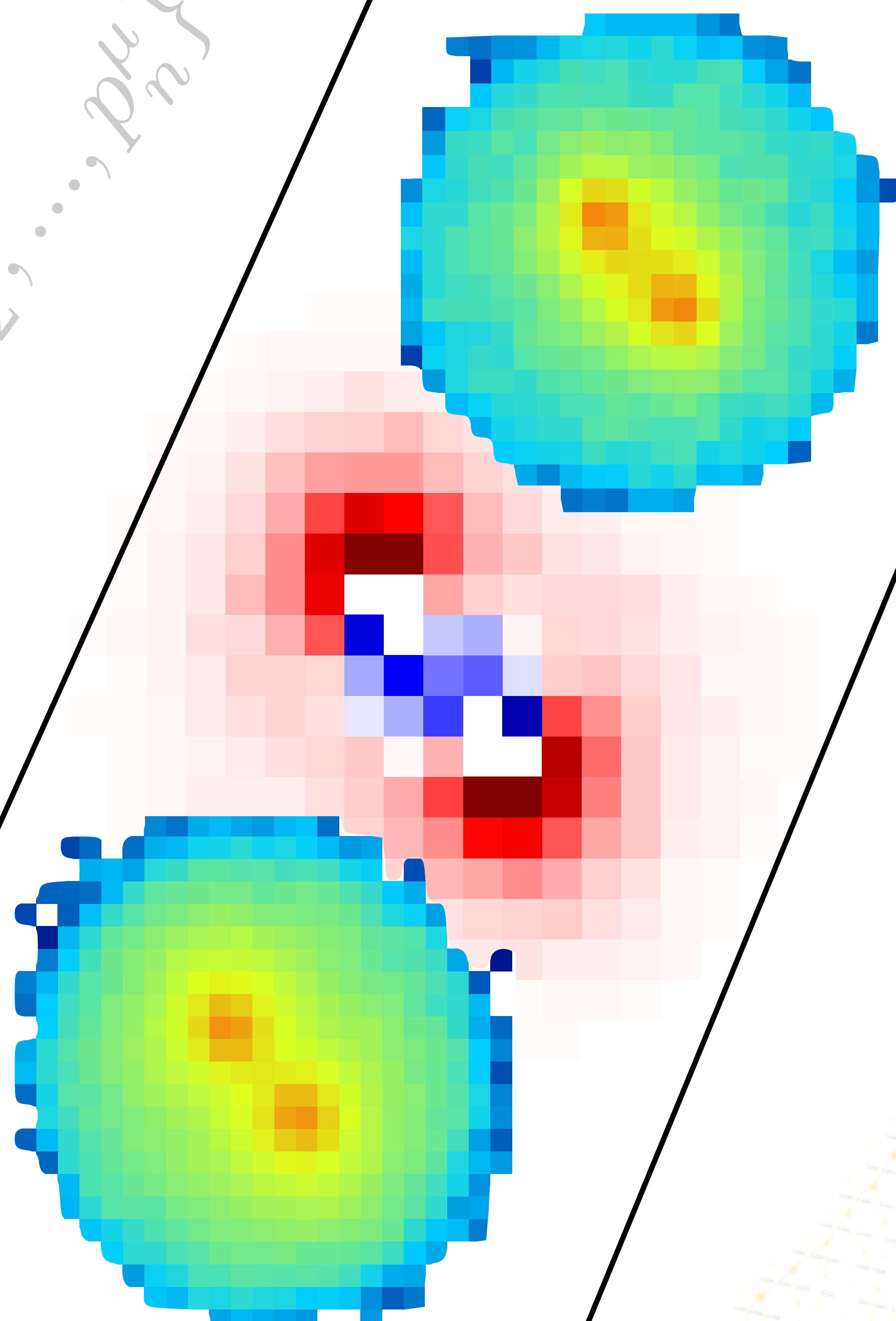


Organizing approaches: how do you think of jets?

Fixed sets

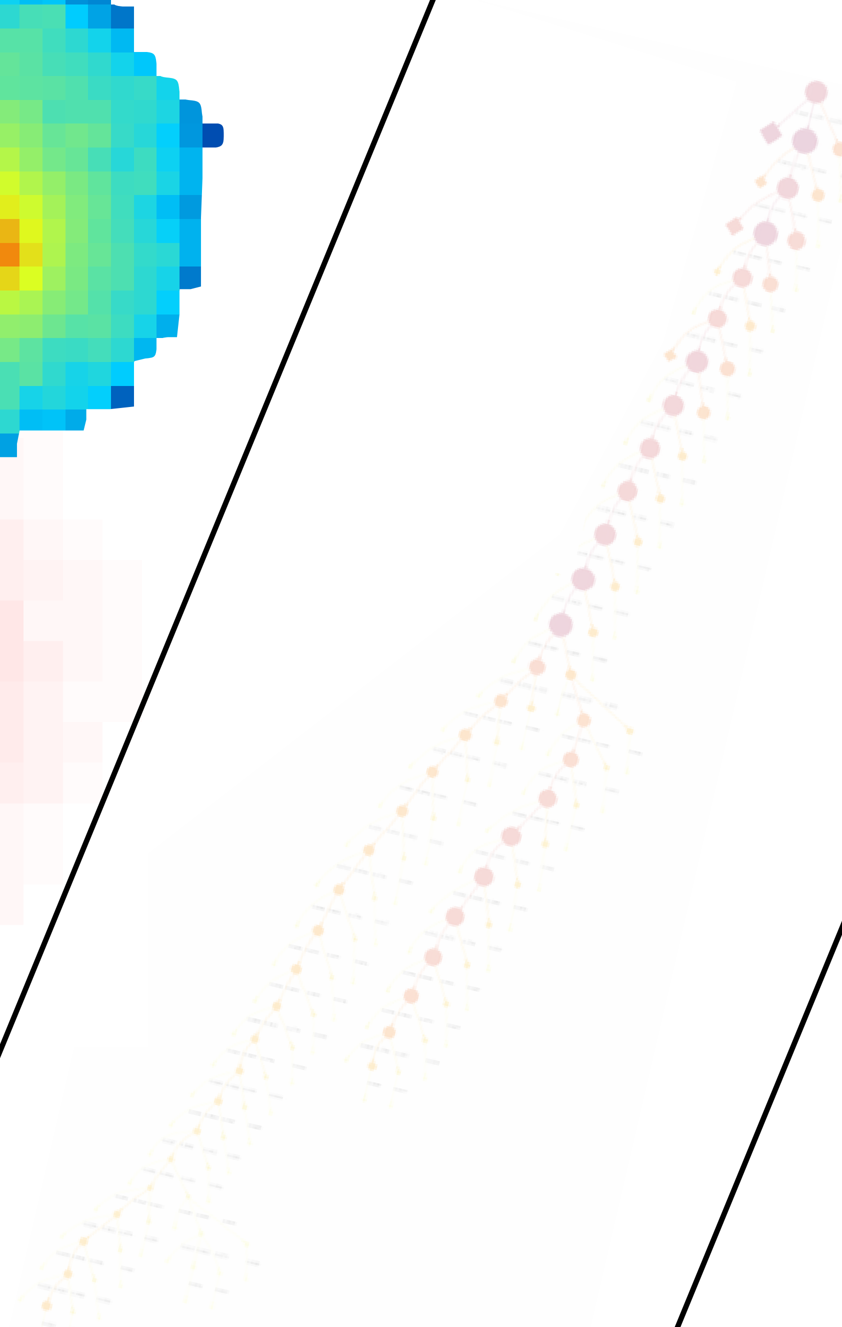
$J = \{p_1^\mu, p_2^\mu, \ldots, p_n^\mu\}$

Images



Jet Images: [1407.5675]

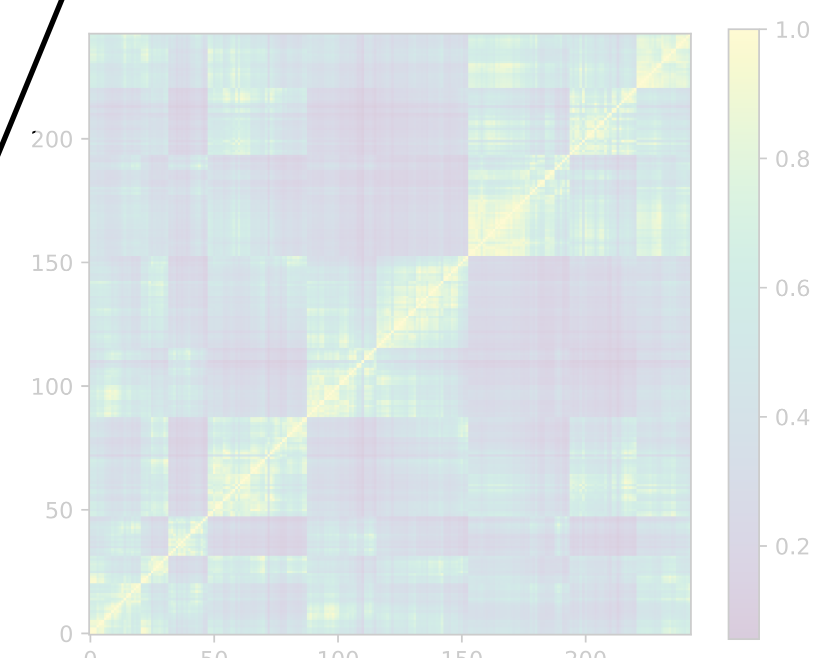
Sequences/ Trees



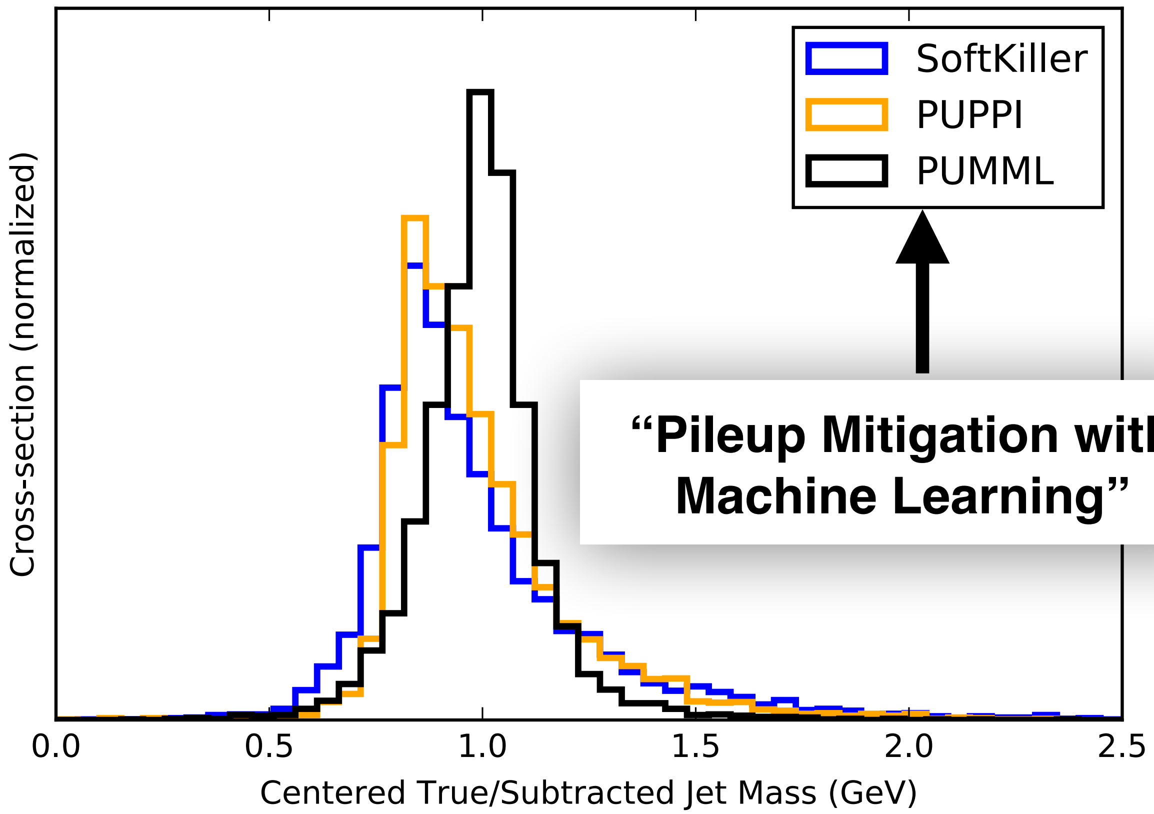
[1702.00748]

Graphs

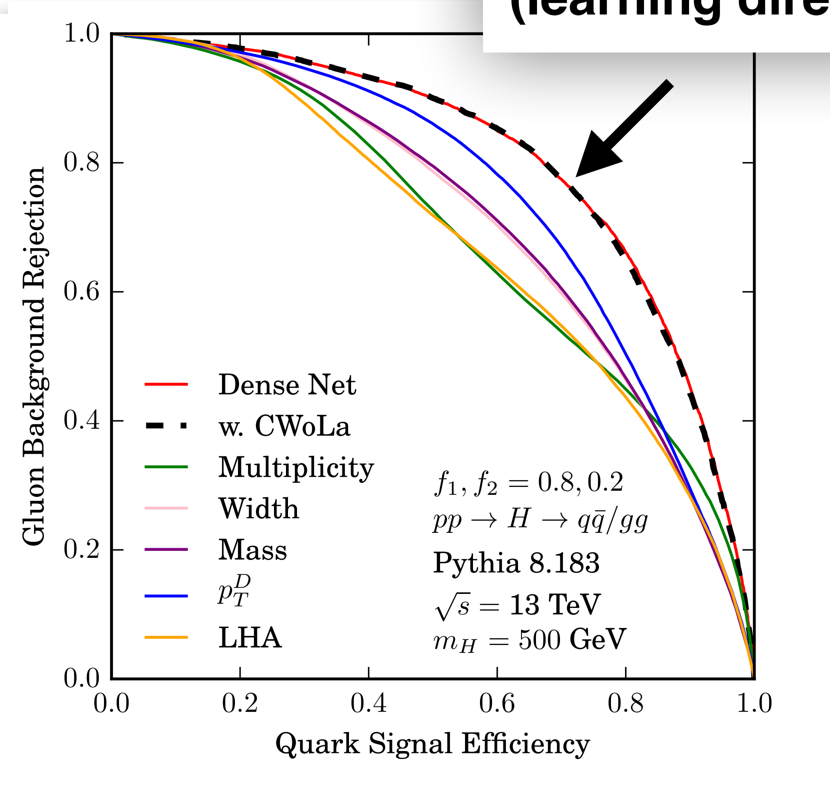
[NIPS DLPS]



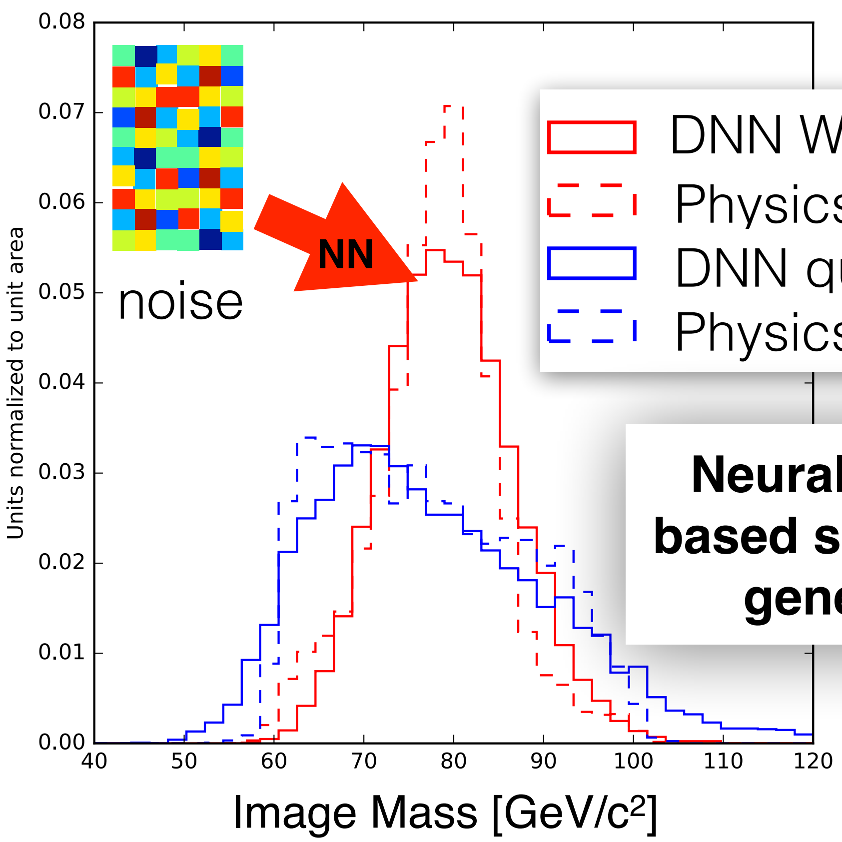
Solving challenges for jets with ML



“Pileup Mitigation with Machine Learning”



Weakly supervised learning (learning directly from data)



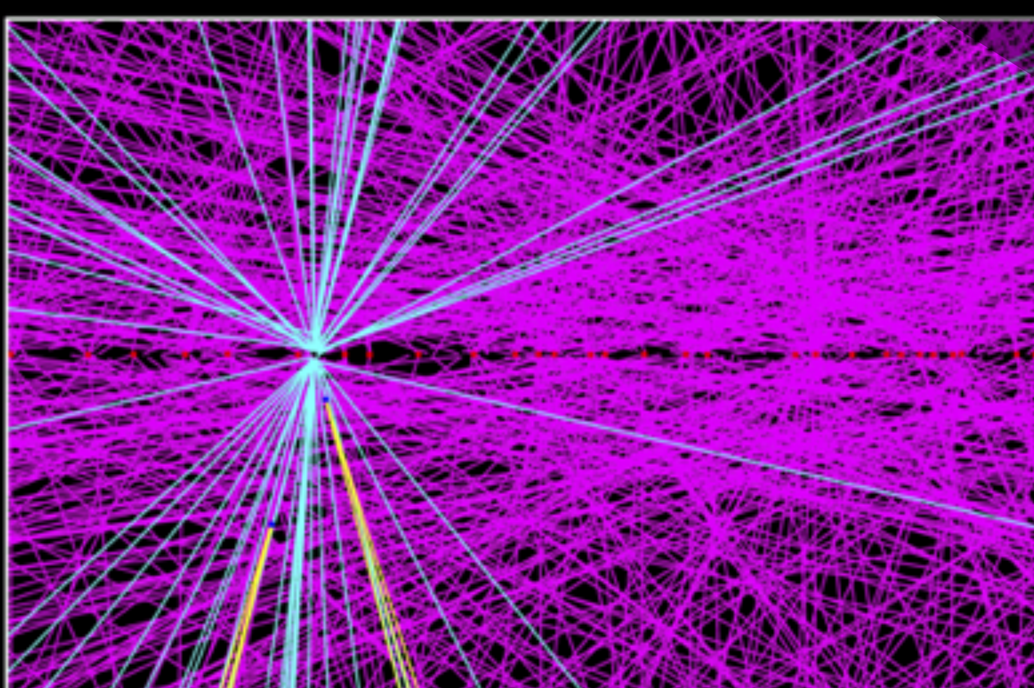
Neural-network based simulation / generation

Machine learning may help us achieve **greater precision**, become more **model independent**, & **learn** about emergent features of the strong force!

Solution 1: Noise ("pileup")

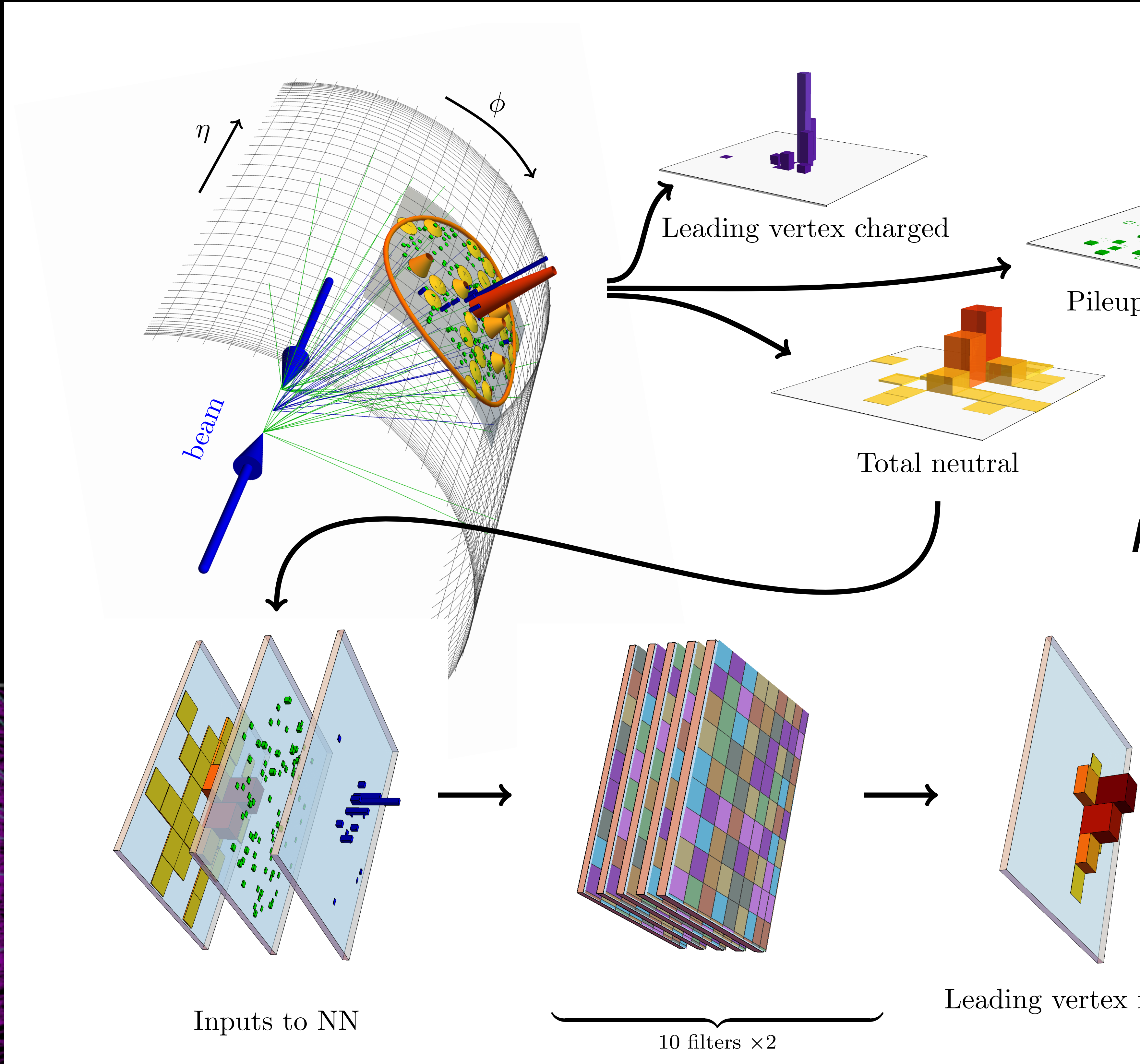
pp collisions at the LHC
don't happen one at a time!

the extra collisions are called **pileup**
and add soft radiation on top of our jets



this is akin to image
de-noising - we can
use ML for that!

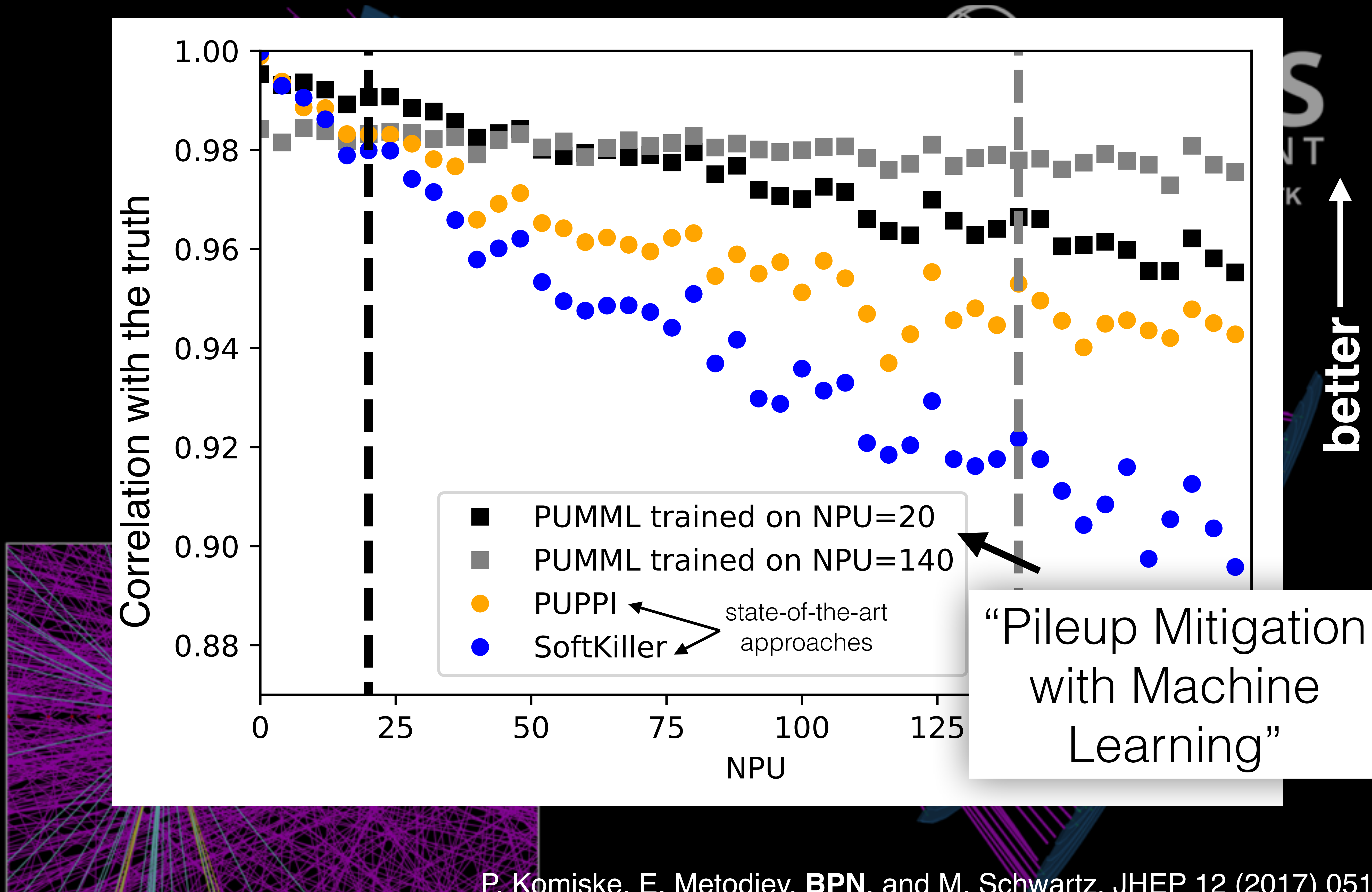
Solution 1: Noise ("pileup")



Strange noise because we can measure ~2/3 of it (charged pileup)

…also a natural application of convolutional NNs!

Solution 1: Noise ("pileup")

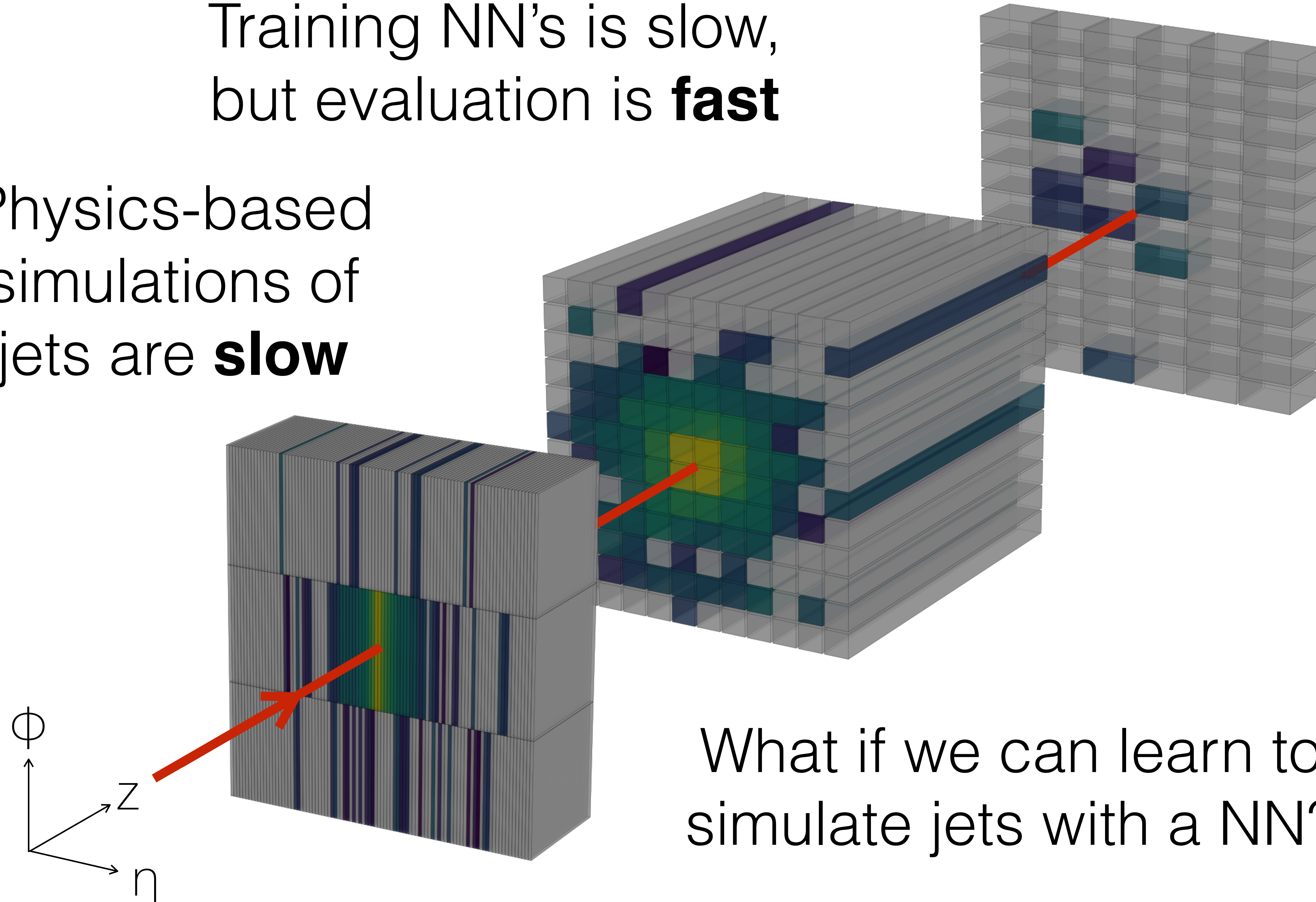


P. Komiske, E. Metodiev, **BPN**, and M. Schwartz, JHEP 12 (2017) 051

Solution 2: Accelerating simulation

Training NN's is slow, but evaluation is **fast**

Physics-based simulations of jets are **slow**

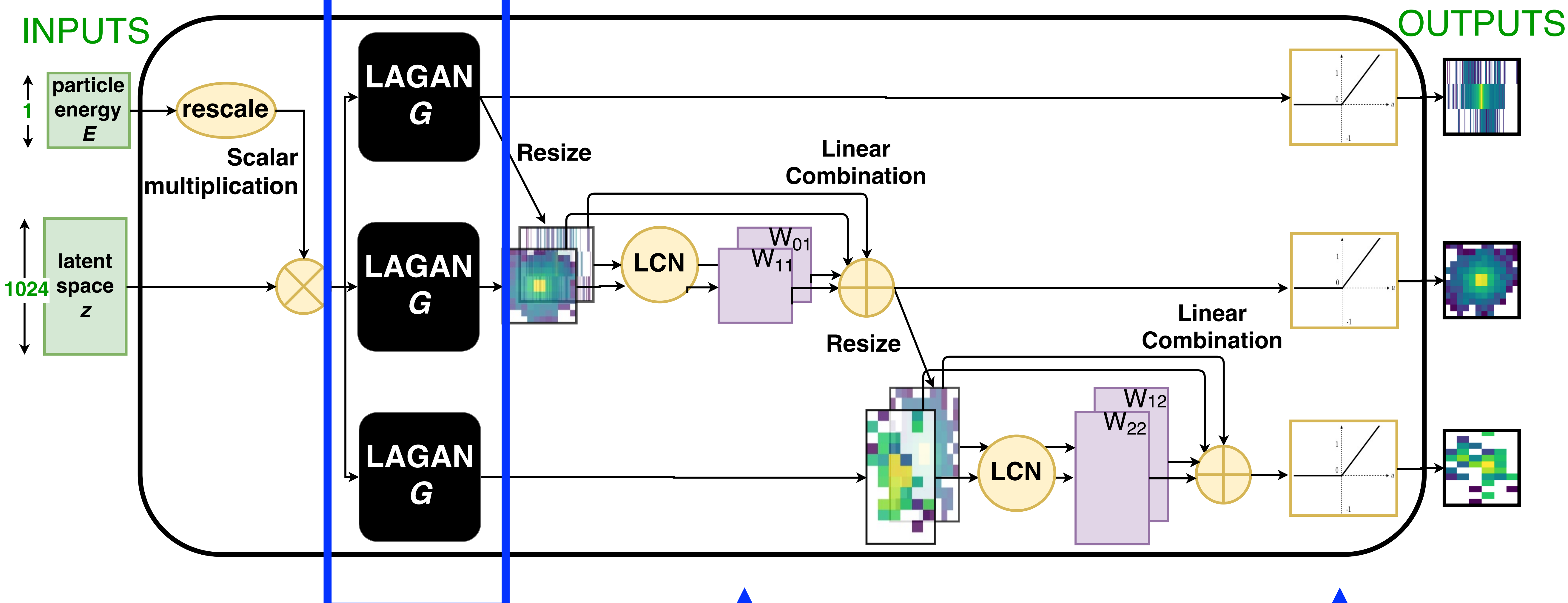


What if we can learn to simulate jets with a NN?

Solution 2: Accelerating simulation

One image per calo layer

One network per particle type; input particle energy

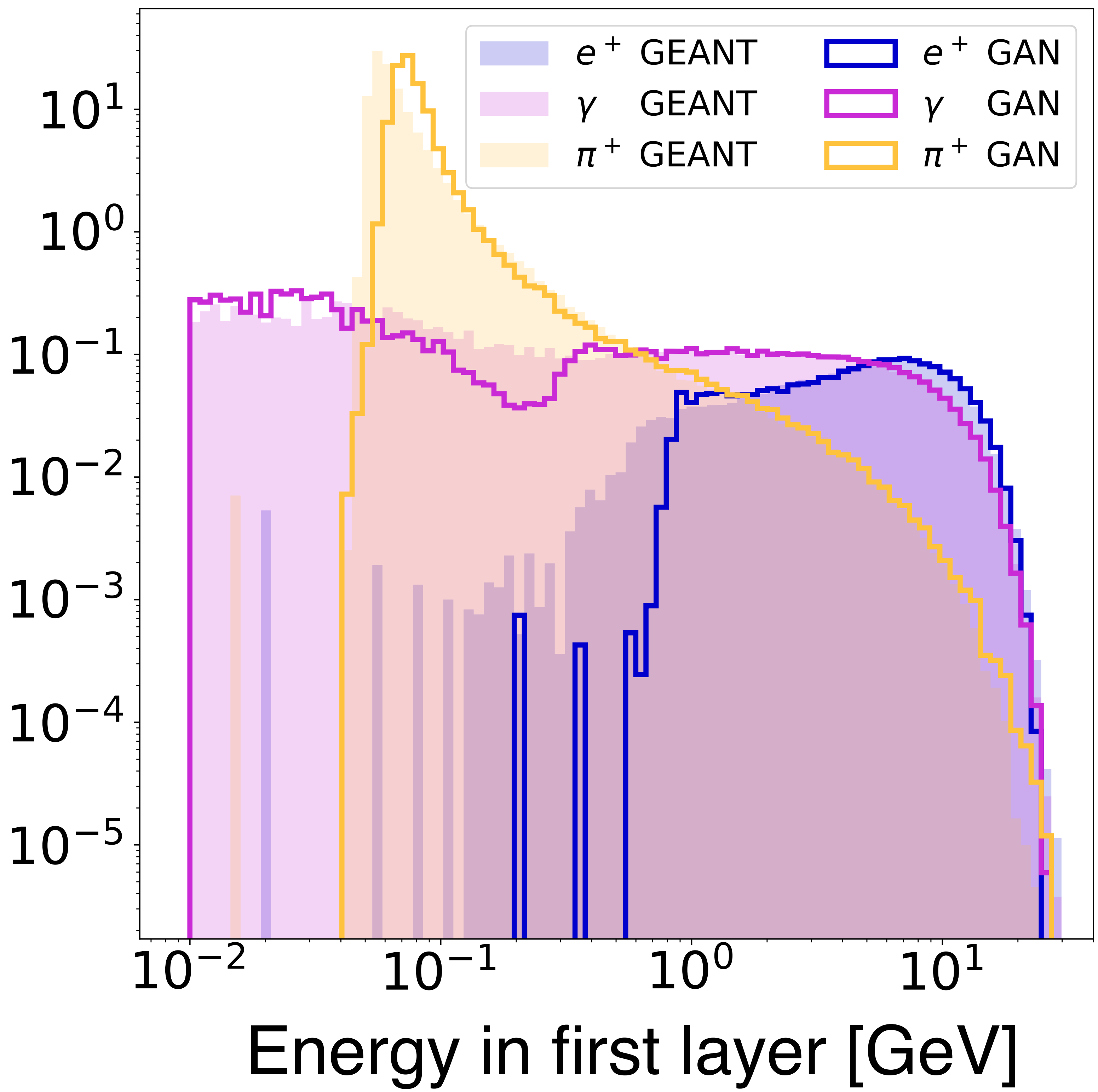


Generator inside Generative Adversarial Network (GAN)

use layer i as input to layer i+1

ReLU to encourage sparsity

Solution 2: Accelerating simulation



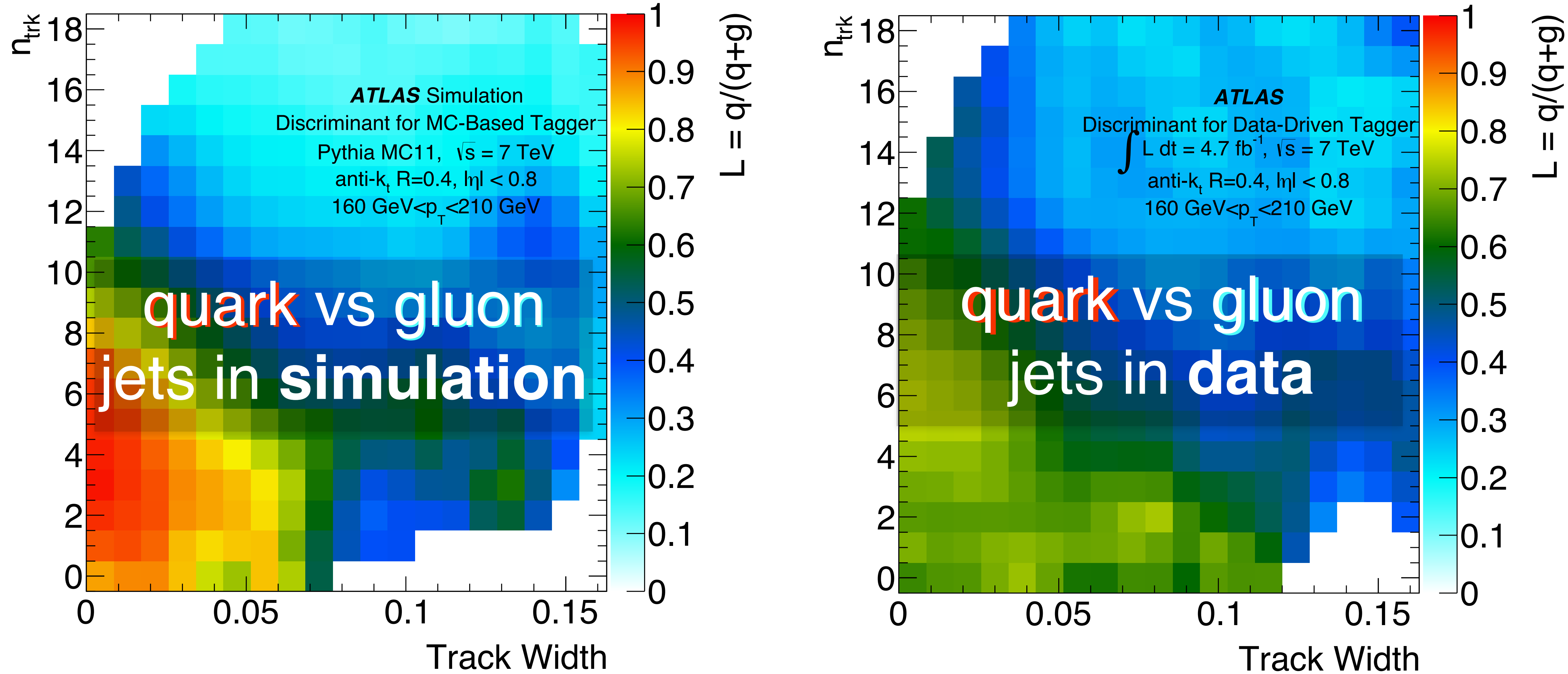
Qualitative agreement; clearly also room for improvement.

Active development at LHC to implement GAN-like approaches (see e.g. this)

up to x10^5 faster!

Solution 3: Learning directly from data

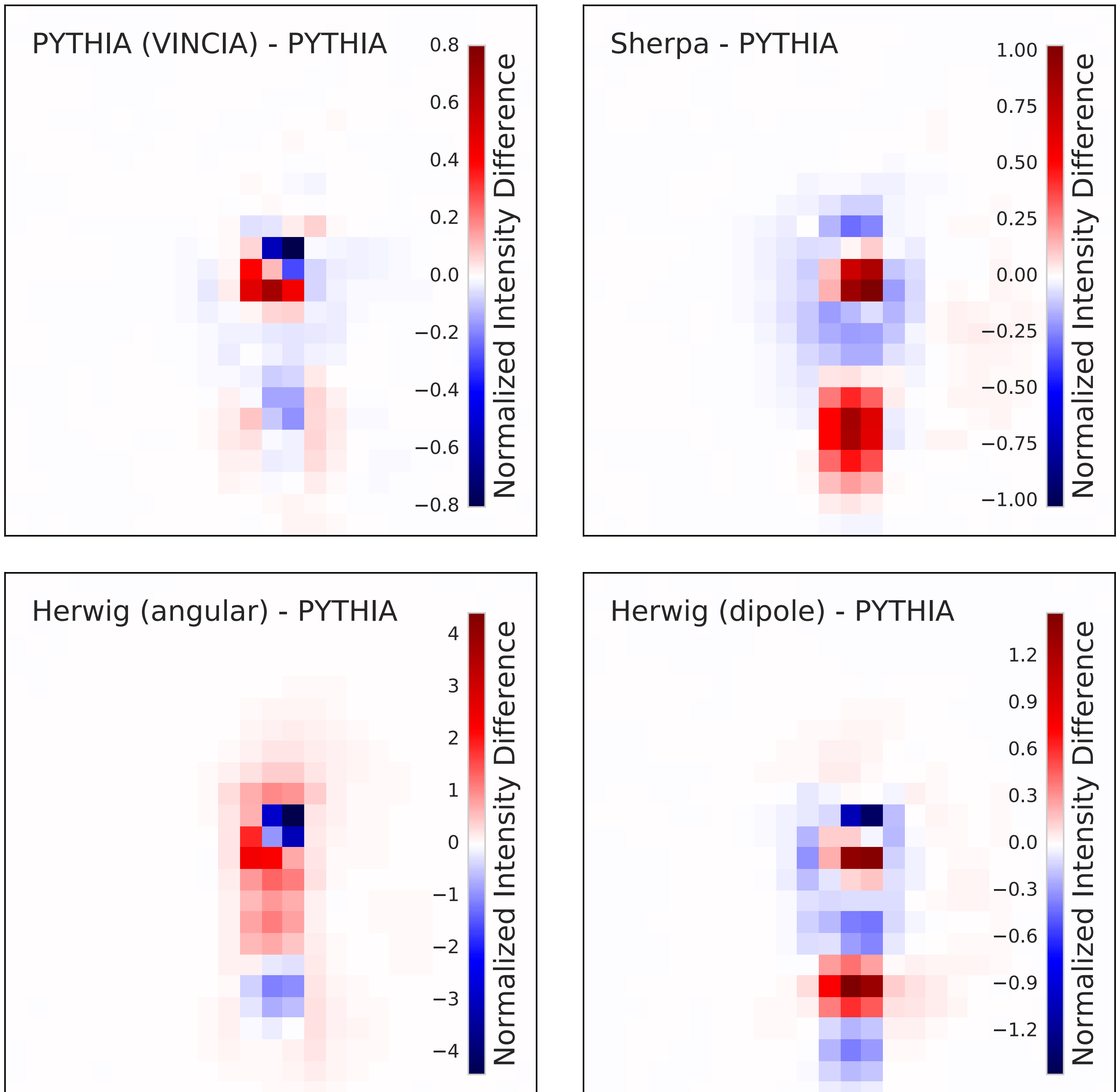
For supervised learning, we depend on labels
labels usually come from simulation



What if data and simulation are very different?
…your classifier will be sub-optimal

Solution 3: Learning directly from data

Boosted Higgs boson jets



J. Barnard et al.
Phys. Rev. D 95, 014018 (2017)

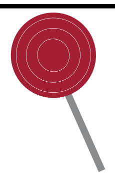
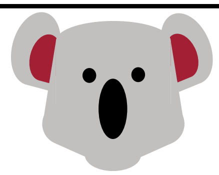
DNN classifiers can **exploit** subtle features

subtle features are **hard to model** !

we need to be careful about which models we use - **only data is correct**

For a mixed approach, see G. Louppe et al.

Two methods

Property	 LLP	 CWoLa
Compatible with any trainable model	✓	✓
No training modifications needed	✗	✓
Training does not need fractions	✗	✓
Smooth limit to full supervision	✗	✓
Works for > 2 mixed samples	✓	?

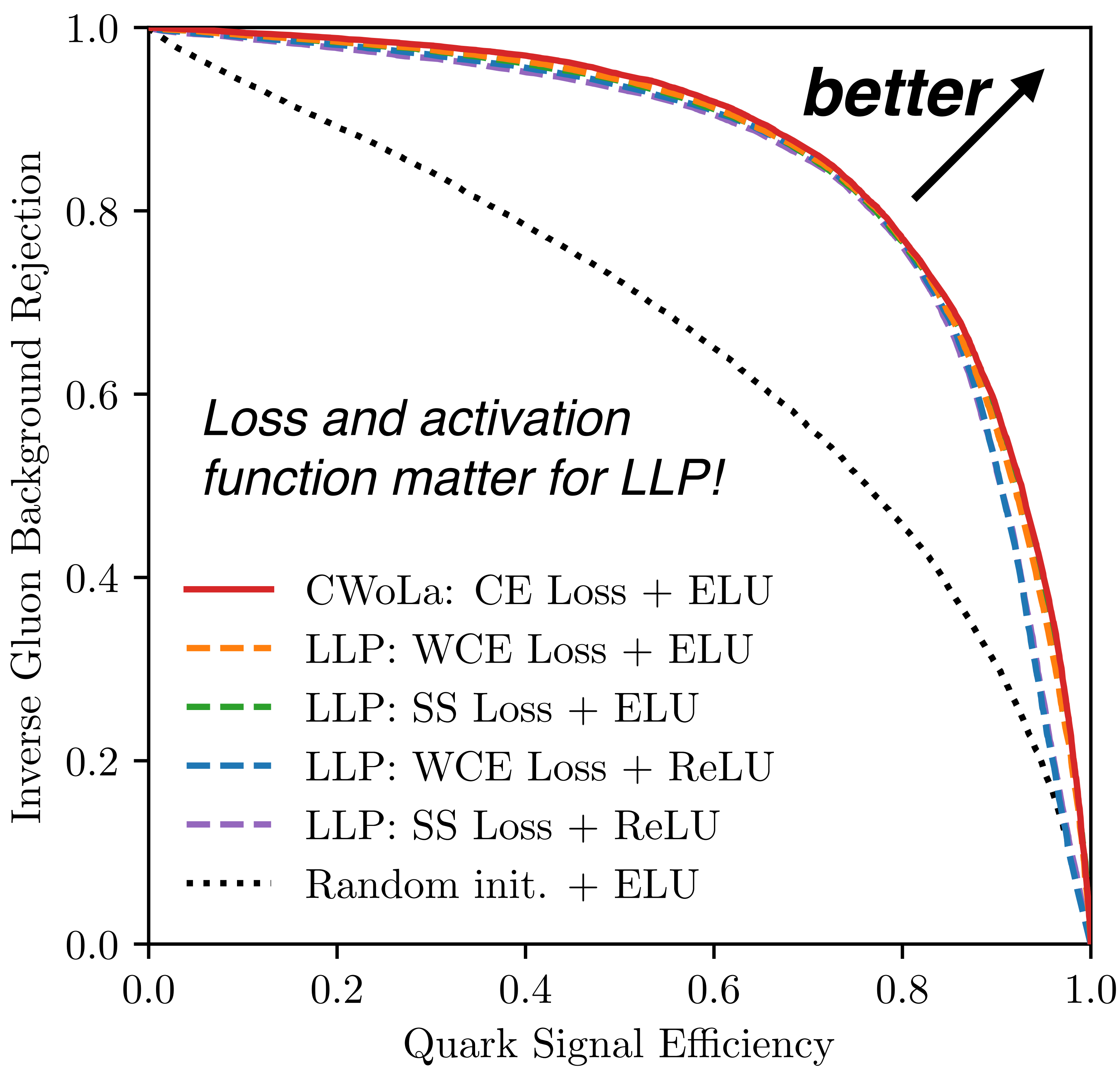
Learning from **L**abel **P**roportions

Classification **w**ith**o**ut **La**bels

L. Dery, **BPN**, F. Rubbo, A. Schwartzman, JHEP 05 (2017) 145

E. Metodiev, **BPN**, J. Thaler, JHEP 10 (2017) 51

Application to jet image classification



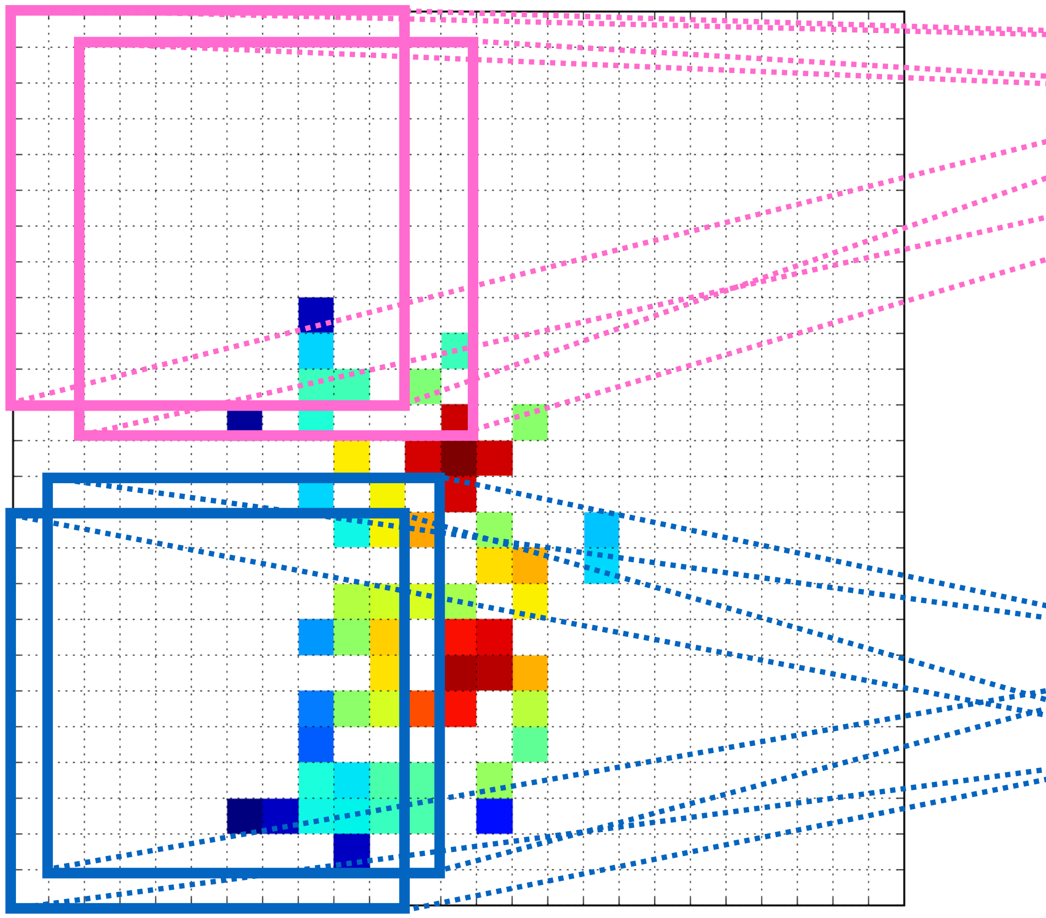
Both methods work well; a promising new direction with many potential applications!

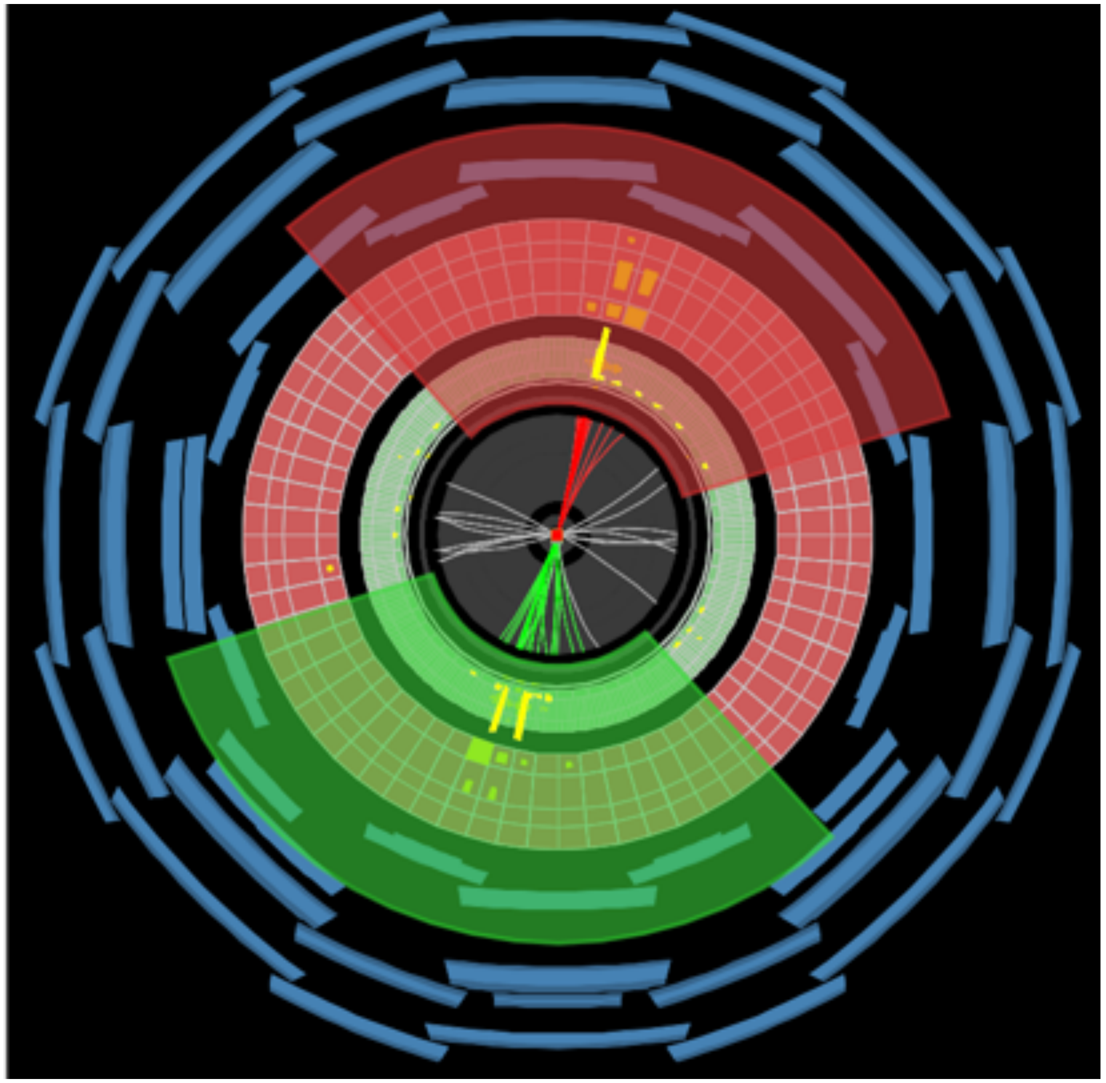
…already ideas for applying this in the context of anomaly detection in this paper.

P. Komiske, E. Metodiev, **BPN**, M. Schwartz: Phys. Rev. D 98, 011502(R)

Conclusions and Outlook

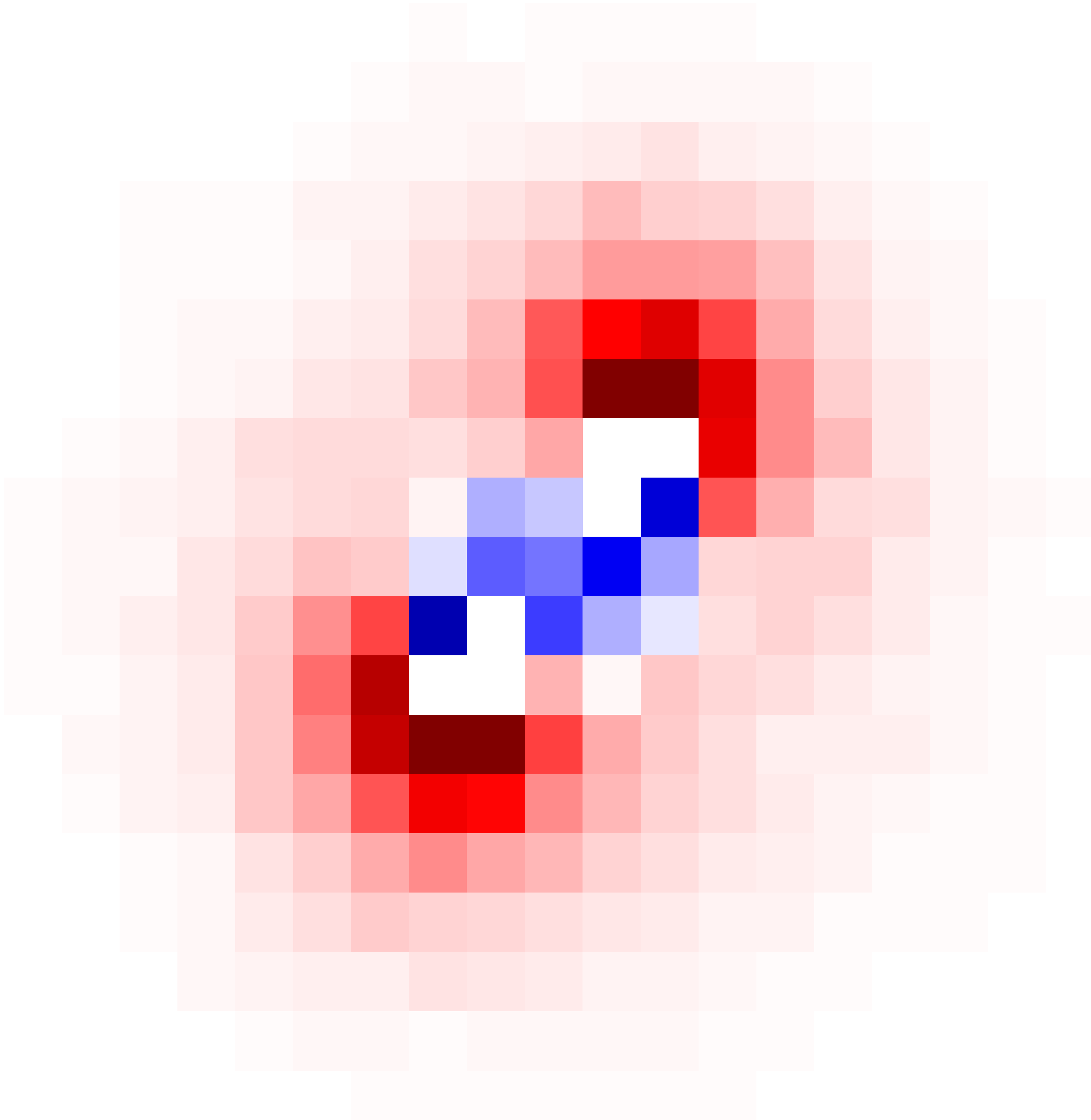
(Jet) image-based NN classification, regression, and generation are powerful tools for fully exploiting the physics program at the LHC





This is only a taste - ML4HEP is a very active field…

…that may hopefully help us understand something new and fundamental about nature!



Fin.